Assembling genome-wide transporter system annotations

Oscar Dias[1], Miguel Rocha[2], Eugénio Ferreira[1], Isabel Rocha[1]

[1]IBB-Institute for Biotechnology and Bioengineering / Centre of Biological Engineering, University of Minho, Campus de Gualtar, 4710-057 Braga, Portugal

[2] Department of Informatics / CCTC, University of Minho, Campus de Gualtar, 4710-057 Braga – PORTUGAL

Abstract

Over the years, several genome-scale metabolic models have been released with compartmentation information. Models such as iMH805/775 [1] (15 compartments) and iMM904 [2] (8 compartments) for Saccharomyces cerevisiae or iRS1563 [3] for Zea mays (6 compartments) include reactions performed in specific cellular organelles, such as mitochondria, chloroplasts (in photosynthetic organisms), lysosomes, cell nucleus or the Golgi apparatus, etc. Thus, cells have specific structures, the transport systems, to assist on the metabolites relocation.

Cellular transport systems are described in databases such as TCDB (http://www.tcdb.org/) maintained by the Saier Lab Bioinformatics Group, the TransportDB (http://www.membranetransport.org/) or the YTPdb (http://homes.esat.kuleuven.be/~sbrohee/ytpdb/index.php/Main_Page) yeast transport protein database. TCDB proposed a classification system, analogous to the Enzyme Classification System (EC Number) [4] though including phylogenetic information, for transport proteins. The phylogenetic information included in the proposed nomenclature is very restrictive as it assigns a distinct TC Number to each transport protein identified in this database. Hence, in this work we propose a system to detect and classify potential transport proteins for a given genome.

A TMHMM (http://www.cbs.dtu.dk/services/TMHMM/) [5] search was performed to identify potential transmembrane domains in the genome. Next, Smith-Waterman (SW) [6] alignments were performed on the transport candidate genes, against the TCDB database [7], to identify sequences encoding proteins with sequences similarities to the known transport systems.

Instead of assigning a similar TC number to the transport system, the classification of the metabolites is proposed, according to the frequency of the metabolite in the SW search and the taxonomy resemblance of each hit organism to the case study organism, and not the transporter systems. The membrane where the transporter is located is predicted by the WoLF PSORT (http://wolfpsort.org/) [8] software.

References

1. Herrgård MJ, Swainston N, Dobson P, Dunn WB, Arga KY, Arvas M, Blüthgen N, Borger S, Costenoble R, Heinemann M, Hucka M, Le Novère N, Li P, Liebermeister W, Mo ML, Oliveira AP, Petranovic D, Pettifer S, Simeonidis E, Smallbone K, Spasić I, Weichart D, Brent R, Broomhead DS, Westerhoff HV, Kirdar B, Penttilä M, Klipp E, Palsson BØ, Sauer U, Oliver SG, Mendes P, Nielsen J, Kell DB: A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. Nature Biotechnology 2008, 26:1155-1160.
2. Mo ML, Palsson BØ, Herrgård MJ: Connecting extracellular metabolomic measurements to intracellular flux states in yeast. BMC systems biology 2009, 3:37.
3. Saha R, Suthers PF, Maranas CD: Zea mays iRS1563: A Comprehensive Genome-Scale Metabolic Reconstruction of Maize Metabolism. PLoS ONE 2011, 6:e21784.
4. Barrett AJ, Canter CR, Liebecq C, Moss GP, Saenger W, Sharon N, Tipton KF, Vnetianer P, Vliegenthart VFG: Enzyme Nomenclature. San Diego: Academic Press; 1992:862.
5. Krogh A, Larsson B, von Heijne G, Sonnhammer EL: Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. Journal of molecular biology 2001, 305:567-80.
6. Smith TF, Waterman MS: Identification of common molecular subsequences. Journal of molecular biology 1981, 147:195-7.
7. Saier MH: A functional-phylogenetic classification system for transmembrane solute transporters. Microbiology and molecular biology reviews : MMBR 2000, 64:354-411.
8. Horton P, Park K-J, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, Nakai K: WoLF PSORT: protein localization predictor. NUCLEIC ACIDS RESEARCH 2007, 35:W585-W587.